

Hölder and Lipschitz Stability of Solution Sets in Programs with Probabilistic Constraints

René Henrion

and

Werner Römisch

Weierstrass Institute for

Humboldt-University Berlin

Applied Analysis and Stochastics

Institute of Mathematics

D-10117 Berlin, Germany

10099 Berlin, Germany

Abstract

We study perturbations of a stochastic program with a probabilistic constraint and r -concave original probability distribution. First we improve our earlier results substantially and provide conditions implying Hölder continuity properties of the solution sets w.r.t. the Kolmogorov distance of probability distributions. Secondly, we derive an upper Lipschitz continuity property for solution sets under more restrictive conditions on the original program and on the perturbed probability measures. The latter analysis applies to linear-quadratic models and is based on work by Bonnans and Shapiro. The stability results are illustrated by numerical tests showing the different asymptotic behaviour of parametric and nonparametric estimates in a program with a normal probabilistic constraint.

1 Introduction

We consider the following optimization problem with chance constraints:

$$(P) \quad \min\{g(x) \mid x \in X, \mathbb{P}(\xi \leq h(x)) \geq p\}.$$

Here ξ is an s -dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is an objective, $X \subseteq \mathbb{R}^m$ is some abstract constraint set, $h : \mathbb{R}^m \rightarrow \mathbb{R}^s$ defines a system of inequalities and $p \in (0, 1)$ is some probability level. The meaning of the probabilistic constraint above is that the system of inequalities $\xi \leq h(x)$ has to be satisfied with probability p at least. The most prominent representative of (P) is given by linear constraints, i.e. $h(x) = Ax$ for some matrix A . By $\mu := \mathbb{P} \circ \xi^{-1} \in \mathcal{P}(\mathbb{R}^s)$ (the space of Borel probability measures on \mathbb{R}^s) we denote the probability distribution of ξ . Throughout the paper we shall make the following basic convexity assumptions:

(BCA) g is convex, X is closed and convex, h has concave components and the probability measure μ is r -concave for some $r < 0$.

The latter property means that μ^r is a convex set function, i.e.,

$$\mu^r(\lambda A + (1 - \lambda)B) \leq \lambda \mu^r(A) + (1 - \lambda) \mu^r(B)$$

holds true for all $\lambda \in [0, 1]$ and for all Borel measurable and convex $A, B \subseteq \mathbb{R}^s$ such that $\lambda A + (1 - \lambda)B$ is Borel measurable too. Note that many prominent multivariate distributions (e.g. normal, Pareto, Dirichlet or uniform distribution on convex, compact sets) share the property of being r -concave for some $r < 0$ (see [12]).

Introducing the distribution function of the probability measure μ as $F_\mu(y) = \mu(z \in \mathbb{R}^s | z \leq y)$, the problem (P) can be equivalently rewritten as

$$(P) \quad \min\{g(x) \mid x \in X, F_\mu(h(x)) \geq p\}.$$

Usually, only partial information about μ is available, and (P) is solved on the basis of some estimation $\nu \in \mathcal{P}(\mathbb{R}^s)$ of μ . Typically, ν is chosen as a parametric or nonparametric estimator of μ . Hence, rather than the original program (P) , some substitute

$$(P_\nu) \quad \min\{g(x) \mid x \in X, F_\nu(h(x)) \geq p\}$$

is solved. Although, at least in principle, arbitrarily good approximations ν of μ can be obtained, it is by no means obvious that the solutions of (P_ν) will well approximate those of $(P = P_\mu)$ as ν tends to μ . Consider, for instance, the following

Example 1 *Let $m = 2$, $s = 2$ and*

$$\min\{x_2 - x_1 \mid x_1 + x_2 \leq 3/2, \mathbb{P}(\xi_1 \leq x_1, \xi_2 \leq x_2) \geq 1/4\},$$

where $\xi = (\xi_1, \xi_2)$ is assumed to be uniformly distributed over the triangle

$$\text{conv}\{(1, 0), 0, 1), (1, 1)\}.$$

Clearly, our basic convexity assumptions (BCA) is satisfied. The (unique) solution of this problem is $(1, 1/2)$. On the other hand, it is easy to construct approximating sequences ν_n of $\mu = \mathbb{P} \circ \xi^{-1}$ such that the substituted problems (P_ν) have no solution for all n or a constant solution $(3/4, 3/4)$ which certainly does not converge to the solution of the original problem.

Although the original data are supposed to be convex, we do not make any assumptions on the data of the perturbed problems (P_ν) . This allows to admit the important class of empirical approximations which lack any convexity or smoothness properties. Since, in general, the solutions of (P_ν) are not unique under the assumptions (BCA), we have to deal with solution sets. The dependence of solutions and optimal values on the parameter ν is described by the set-valued mapping $\Psi : \mathcal{P}(\mathbb{R}^s) \rightrightarrows \mathbb{R}^m$ and the extended-valued function $\varphi : \mathcal{P}(\mathbb{R}^s) \rightarrow \bar{\mathbb{R}}$ via

$$\begin{aligned} \Psi(\nu) &= \text{argmin}\{g(x) \mid x \in X, F_\nu(h(x)) \geq p\} \\ \varphi(\nu) &= \inf\{g(x) \mid x \in X, F_\nu(h(x)) \geq p\}. \end{aligned}$$

We are interested in conditions formulated for the data of the original problem (P) such that Ψ and φ behave stable locally around the fixed measure μ . In order to measure

distances among parameters and among solutions, we rely on the Kolmogorov metric between probability measures

$$d_K(\nu_1, \nu_2) = \sup_{z \in \mathbb{R}^s} |F_{\nu_1}(z) - F_{\nu_2}(z)| \quad (\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^s)).$$

and on the Hausdorff distance between closed subsets of \mathbb{R}^m

$$d_H(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\} \quad (A, B \subseteq \mathbb{R}^m).$$

Qualitative stability results in the sense of $d_H(\Psi(\mu), \Psi(\nu)) \rightarrow 0$ as $d_K(\mu, \nu) \rightarrow 0$ have been obtained in [5] based on earlier works like [14] and [6]. These results guarantee that, under the imposed conditions (see Theorem 2 below), cluster points of approximating solutions will be solutions of the original problem and that any solution of the original problem is the limit of a sequence of approximating solutions. For further work in this direction we refer to [4, 9, 11, 15, 16].

Beyond qualitative stability it is of much interest to know how fast solutions or optimal values of approximating problems converge to original solutions, which is a question of quantitative stability. Recall that Ψ is Hausdorff-Hölder continuous with rate $\kappa > 0$ at μ , if there are $L, \delta > 0$ such that

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L [d_K(\mu, \nu)]^\kappa \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \delta. \quad (1)$$

There exists an immediate link between Hausdorff-Hölder continuity with rate κ of of the solution set mapping and exponential bounds for empirical solution estimates: Consider independent s -dimensional random vectors ξ_1, \dots, ξ_N on $(\Omega, \mathcal{A}, \mathbb{P})$ having common distribution μ . Then, $\mu_N(\omega) := N^{-1} \sum_{i=1}^N \delta_{\xi_i(\omega)}$ (with δ_z denoting the Dirac measure placing mass one at $z \in \mathbb{R}^s$) is an empirical measure approximating μ as $N \rightarrow \infty$. The deviation between the original solution set and its empirical approximation can be estimated as follows (see Proposition 6 in [6]):

$$\begin{aligned} & \exists C > 0 \forall N \in \mathbb{N} \forall \varepsilon > 0 \\ P(d_H(\Psi(\mu), \Psi(\mu_N)) \geq \varepsilon) & \leq C [N \cdot \lambda(\varepsilon, \delta, \kappa, L)]^{s-0.5} \exp(-2N \cdot \lambda(\varepsilon, \delta, \kappa, L)), \end{aligned} \quad (2)$$

where $\lambda(\varepsilon, \delta, \kappa, L) = [\min\{\delta, (\varepsilon/L)^{1/\kappa}\}]^2$ and δ, L, κ refer to (1).

Conditions for Hausdorff-Hölder continuity of Ψ at rate $\kappa = 1/2$ were obtained in [6] for the special case of linear chance constraints with convex-quadratic objective and in [7] for the more general setting of the above data assumptions (BCA). The first part of this paper is devoted to a substantial improvement of the previous results in two directions: firstly, the mentioned results relate to so-called localized solution sets rather than to the solution sets themselves. This technical restriction seemed to be a necessary consequence of considering non-convex perturbations of the original convex measure. It turns out, however, that one can exploit additional arguments provided in [5] in order to get rid of localizations. Of course, statements on stability of solution sets themselves as in (1) are easier to interpret than their localized counterparts. Secondly, all previous results on quantitative stability of Ψ essentially relied on the condition

$$\Psi(\mu) \cap \operatorname{argmin}\{g(x) \mid x \in X\} = \emptyset, \quad (3)$$

which means that no solution of (P) is a solution of the relaxed problem with the chance constraint removed and vice versa. In this paper we shall obtain the same results without requiring such kind of strict complementarity condition.

Specializing our setting to linear chance constraints, the best (largest) rate we can obtain is $\kappa = 1/2$ provided that the random variable has at least dimension $s = 2$. Of course, the exponential bound in (2) improves with increasing κ . Thus, it is of much interest to find conditions ensuring even Hausdorff-Lipschitz continuity of Ψ ($\kappa = 1$). It is interesting to note that a Lipschitz rate results for linear chance constraints with 1-dimensional random variable ξ (but with arbitrary dimension of the decision variable x). Yet, this observation seems to be too restrictive for practical relevance. The second part of the paper investigates more reasonable settings and conditions for Lipschitz rates. Two basic additional requirements turn out to be crucial then: firstly the approximating measures ν can no longer be arbitrary but have to be restricted to class \mathcal{C}^1 in a sense to be precised. Secondly, the strict complementarity condition (3) which was dispensable for the Hölder rate $\kappa = 1/2$, has to be incorporated into the set of conditions now.

2 Hölder Stability

The main result of this section is stated with the technical details of proof left to the appendix. We start by recalling two results which are needed for the proof of our main theorem.

Theorem 2 (see [5], Th. 1) *In addition to the basic convexity assumptions (BCA), let the following conditions be satisfied at the fixed probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$:*

1. $\Psi(\mu)$ is nonempty and bounded.
2. There exists some $\hat{x} \in X$ such that $F_\mu(h(\hat{x})) > p$.

Then, $\Psi : \mathcal{P}(\mathbb{R}^s) \rightrightarrows \mathbb{R}^m$ is upper semicontinuous in the sense of Berge at μ , and there exist constants $L, \delta > 0$, such that

$$\Psi(\nu) \neq \emptyset \text{ and } |\varphi(\nu) - \varphi(\mu)| \leq Ld_K(\nu, \mu) \text{ for all } \nu \in \mathcal{P}(\mathbb{R}^s) \text{ with } d_K(\nu, \mu) < \delta.$$

The second result provides a useful two-level decomposition of the parametric program (P_ν) :

Lemma 3 (see [5], Lemma 1) *Under the assumptions of Theorem 2 let V be an open ball containing $\Psi(\mu)$. Set*

$$\begin{aligned} Y_V &= [h(X \cap \text{cl } V) + \mathbb{R}_-^s] \cap F_\mu^{-1}([p/2, 1]) \\ \pi(y) &= \inf\{g(x) \mid x \in X \cap \text{cl } V, h(x) \geq y\}, \\ Y(\nu) &= \operatorname{argmin}\{\pi(y) \mid y \in Y_V, F_\nu(y) \geq p\} \quad (\nu \in \mathcal{P}(\mathbb{R}^s)) \\ \sigma(y) &= \operatorname{argmin}\{g(x) \mid x \in X \cap \text{cl } V, h(x) \geq y\} \quad (y \in Y_V). \end{aligned}$$

Then it holds that

1. Y_V is convex and compact.
2. π is convex, finite and lower semicontinuous on Y_V .
3. There is some $\delta > 0$ such that for all $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \delta$

$$\varphi(\nu) = \inf\{\pi(y) \mid y \in Y_V, F_\nu(y) \geq p\} \quad (4)$$

$$\Psi(\nu) = \sigma(Y(\nu)) \quad (5)$$

4. $Y : \mathcal{P}(\mathbb{R}^s) \rightrightarrows \mathbb{R}^s$ is upper semicontinuous at μ .

Now, we are in a position to state the main result of this section. We refer to the notation of Lemma 3.

Theorem 4 *In addition to the basic convexity assumptions (BCA), let the following conditions be satisfied at some fixed $\mu \in \mathcal{P}(\mathbb{R}^s)$:*

1. $\Psi(\mu)$ is nonempty and bounded.
2. There exists some $\hat{x} \in X$ such that $F_\mu(h(\hat{x})) > p$.
3. F_μ^r is strongly convex on some convex open neighbourhood U of $Y(\mu)$, where $r < 0$ is chosen from (BCA) such that μ is r -concave.
4. σ is Hausdorff Hölder continuous with rate κ^{-1} on Y_V .

Then, Ψ is Hausdorff Hölder continuous with rate $(2\kappa)^{-1}$ at μ , i.e., there are $L, \delta > 0$ such that

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L [d_K(\mu, \nu)]^{1/(2\kappa)} \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \delta.$$

Proof.

Combine Lemma 13 with Proposition 12. □

The first assumption of Theorem 4 is of technical nature. It can be enforced, for instance, by compactness of X (the nonemptiness of the compact constraint set is then a consequence of the second assumption). The second assumption can be interpreted as a Slater condition (see proof of Lemma 13). In special situations, its verification is possible without explicit knowledge of the measure μ . For instance, in the situation of linear chance constraints under nonnegativity restrictions ($h(x) = Ax, X = \mathbb{R}_+^m$), it suffices to know that $A \geq 0, A \neq 0$. Indeed, for $v := A\mathbf{1}$ with $\mathbf{1} = (1, \dots, 1)$, one has $v \geq 0, v \neq 0$. Consequently, $\lim_{\lambda \rightarrow \infty} F_\mu(\lambda v) = 1$. Since $p < 1$, there is some $\lambda > 0$ such that $F_\mu(\lambda v) > p$. Hence, $F_\mu(A\hat{x}) > p$ for $\hat{x} := \lambda\mathbf{1} \in X$, which is condition 2. in Theorem 4. An alternative situation occurs when $X = \mathbb{R}^m$ and A has linearly independent rows.

The third assumption of Theorem 4 is satisfied for r -concave measures ($r < 0$) for which F_μ^r is strongly convex on bounded, convex sets (because $Y(\mu)$ is compact, see Lemma 3). An example for such measure is the multivariate normal distribution with

independent components (see [8]). But even measures lacking this last property may still satisfy the third assumption. For instance, the uniform distribution over multidimensional rectangles is r -concave for any $r < 0$, such that F_μ^r is strongly convex on this rectangle. All one has to know then is that $Y(\mu)$ is contained in the rectangle too. Unfortunately, not all uniform distributions over polytopes share the required strong convexity property (e.g., the uniform distribution over the triangle $\text{conv}\{(1, 0), (0, 1), (1, 1)\}$ is r -concave for any $r < 0$ but F_μ^r fails even to be strictly convex on this triangle). If h is linear, i.e., $h(x) = Ax$, then the strong convexity assumption can be simplified in the sense that it is supposed to hold on some convex open neighbourhood U of $A(\Psi(\mu))$.

In the last assumption of Theorem 4, it is assumed that some Hölder continuity of σ with respect to the Hausdorff distance is known. This is the case, for instance, for linear mappings h , polyhedral sets X and convex-quadratic functions g . Then the Hölder rate for σ equals 1 (see Th. 4.2 in [10] or Prop. 2.4 in [7]) and we have the following Corollary to Theorem 4:

Corollary 5 *In addition to the basic convexity assumptions (BCA), let g be convex-quadratic, h linear and X polyhedral. Then, supposing the first three assumptions of Theorem 4, Ψ is Hausdorff Hölder continuous with rate $1/2$ at μ , i.e., there are $L, \delta > 0$ such that*

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L\sqrt{d_K(\mu, \nu)} \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \delta.$$

Apart from this application to linear chance constraints defined by h , there is also a chance of identifying a Hölder rate of σ in the more general situation considered here, when h has concave components (see Prop. 2.4 in [7] for more details).

The following example demonstrates that the Hölder rate obtained in Theorem 4 and in Corollary 5 is sharp.

Example 6 *In problem (P), let $m = s = 2$, $X = \mathbb{R}^2$, $h(x) = x$, $g(x) = (x_1 + x_2 - \sqrt{2})^2$, $p = 0.5$ and $\mu =$ uniform distribution over the unit square $[0, 1]^2$. Evidently, these data satisfy all the basic assumptions formulated in the introduction (in particular, μ is log-concave, hence r -concave for any $r < 0$). Next we verify the assumptions of Theorem 4: since the distribution function of μ satisfies $F_\mu(x) = x_1x_2$ for all $(x_1, x_2) \in [0, 1]^2$, it follows that $\Psi(\mu) = \{(\sqrt{1/2}, \sqrt{1/2})\}$ which entails 1. in Theorem 4. It is elementary to verify that one may assume $Y(\mu) \subseteq [0, 1]^2$ (after shrinking the open ball $V \supseteq \Psi(\mu)$ used in Lemma 3). Since F_μ is strongly log-concave on $[0, 1]^2$, F_μ^r is strongly convex for any $r < 0$, whence 3. With $\hat{x} := (1, 1)$, one has $F_\mu(\hat{x}) = 1 > p$, which is 2. Finally, since g is convex-quadratic, X is trivially a polyhedral set and h is linear, it follows that σ is Hausdorff Lipschitz continuous (see remarks above Corollary 5). This provides 4. with $\kappa = 1$, and, thus, Theorem 4 ensures that Ψ is Hausdorff Hölder continuous with rate $1/2$ at μ . This rate is sharp. Indeed, considering the perturbed measures $\nu_\varepsilon \in \mathcal{P}(\mathbb{R}^s)$ defined for $\varepsilon > 0$ as uniform distributions over the rectangles $[-\varepsilon, 1 - \varepsilon]^2$, a straightforward calculation shows that*

$$\Psi(\nu_\varepsilon) = \text{conv}\{(a_\varepsilon, b_\varepsilon), (b_\varepsilon, a_\varepsilon)\} \quad \text{and} \quad d_K(\mu, \nu_\varepsilon) = \varepsilon(1 + \varepsilon),$$

where $a_\varepsilon/b_\varepsilon = \sqrt{1/2} \pm \sqrt{\varepsilon(\sqrt{2} + \varepsilon)}$. Consequently,

$$d_H(\Psi(\mu), \Psi(\nu)) = \sqrt{2}\sqrt{\varepsilon(\sqrt{2} + \varepsilon)} \geq \sqrt{\varepsilon(1 + \varepsilon)} = \sqrt{d_K(\mu, \nu_\varepsilon)},$$

which shows that the Hölder rate $1/2$ cannot be improved in this example.

A more sophisticated counter-example (Example 2.10 in [7]) confirms that the Hölder rates of Theorem 4 and Corollary 5 are sharp even in case of linear objective functions g . This observation is easily modified to construct a counter-example with strongly convex objective function (which was not the case in the example above). On the other hand, all these examples live in \mathbb{R}^2 . The following Theorem confirms that the Hölder rates of Theorem 4 and Corollary 5 can be improved as long as the random variable ξ is one-dimensional (the decision variable x is arbitrary). Moreover, in this special case no strong convexity assumption is needed for the measure μ (condition 3. in Theorem 4):

Theorem 7 *In addition to the basic convexity assumptions (BCA), let $s = 1$ and assume conditions 1., 2. and 4. of Theorem 4. Then, Ψ is Hausdorff Hölder continuous with rate κ^{-1} at μ . In the context of Corollary 5, Ψ is even Hausdorff Lipschitz continuous (rate $\kappa = 1$) at μ .*

Proof.

Combine Lemma 14 with Proposition 12. □

3 Lipschitz Stability

The Lipschitz result of Theorem 7 (in the context of Corollary 5) is based on the one-dimensionality of the random variable which is rather restrictive in stochastic programming. In order to derive Lipschitz stability in a multivariate setting, one has to impose further conditions and also to restrict the class of considered measures (for the original as well as the approximating one). The subsequent analysis relies on general stability results obtained in [1, 2]. The following theorem gives a reduction of those results to the setting which will be of interest here:

Theorem 8 *(see [2], Th. 4.8.1) Consider the parametric optimization problem*

$$\min\{f(x) | G(x, u) \in K\},$$

where $f : \mathbb{R}^m \rightarrow \mathbb{R}$, $G : \mathbb{R}^m \times U \rightarrow \mathbb{R}^q$, U is a Banach space, $K = \mathbb{R}^{q_1} \times \{0\}_{q_2}$, $q_1 + q_2 = q$. Denote by $S(u) := \arg \min\{f(x) | G(x, u) \in K\}$ the parametric solution set and fix some parameter $u_0 \in U$. Let the following conditions hold true:

1. f and G are $\mathcal{C}^{1,1}$ functions (differentiable with Lipschitz continuous derivative).
2. $S(u_0) \neq \emptyset$ and S is uniformly bounded in a neighbourhood of u_0 .

3. f satisfies a second order growth condition with respect to $S(u_0)$, i.e., there exist a neighbourhood \mathcal{V} of $S(u_0)$ and a constant $c > 0$ such that

$$\begin{aligned} f(x) &\geq f_0 + c \text{dist}^2(x, S(u_0)) \quad \forall x \in \mathcal{V}, G(x, u_0) \in K \\ (f_0 &= \min\{f(x) | G(x, u_0) \in K\}). \end{aligned}$$

4. For all $x \in S(u_0)$ it holds that

$$\{\nabla_x G_i(x, u_0)\}_{i=1, \dots, q_2} \cup \{\nabla_x G_j(x, u_0)\}_{j \in I(x)}$$

is a set of linear independent vectors in \mathbb{R}^m , where

$$I(x) = \{j \in \{1, \dots, q_1\} | G_j(x, u_0) = 0\}.$$

Then, S is upper Lipschitz at u_0 , i.e., there are constants $L, \delta > 0$ such that

$$\text{dist}(x, S(u_0)) \leq L \|u - u_0\| \quad \forall x \in S(u) \forall u \in U, \|u - u_0\| < \delta.$$

In order to apply Theorem 8 to our parametric problem (P_ν) , we have to interpret the parameter u as distribution functions F_ν where $\nu \in \mathcal{P}(\mathbb{R}^s)$. However, condition 1. requires to restrict the admissible class of measures to those having $\mathcal{C}^{1,1}$ distribution function. More precisely, we introduce the following space:

$$\mathcal{C}_b^{1,1}(\mathbb{R}^n) := \{f \in \mathcal{C}^1(\mathbb{R}^n) | f \text{ is bounded and has a bounded, Lipschitzian derivative}\}$$

With the norm

$$\|f\|_b^{1,1} := \max \left\{ \sup_{x \in \mathbb{R}^n} |f(x)|, \sup_{x \in \mathbb{R}^n} \|\nabla f(x)\|, \sup_{x, y \in \mathbb{R}^n, x \neq y} \frac{\|\nabla f(x) - \nabla f(y)\|}{\|x - y\|} \right\},$$

$\mathcal{C}_b^{1,1}(\mathbb{R}^n)$ becomes a Banach space.

In the parametric problem (P_ν) , let us specify the general convexity assumptions (BCA) in the following sense:

- The objective function g is convex-quadratic, i.e., $g(x) = \langle x, Hx \rangle + \langle c, x \rangle$ for some positive semidefinite (m, m) -matrix H ($H = 0$ possible) and some $c \in \mathbb{R}^m$.
- $h(x) = Ax$, where A is a matrix of order (s, n) .
- X is a polyhedron and has an explicit description

$$X = \{x \in \mathbb{R}^m | \langle \alpha_j, x \rangle \leq a_j \ (j = 1, \dots, \tilde{q}_1); \ \langle \beta_i, x \rangle = b_i \ (i = 1, \dots, \tilde{q}_2)\}.$$

- For some fixed probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ it holds that μ is r -concave for some $r < 0$.

Now, we are in a position to formulate the desired stability result:

Theorem 9 *Let the following conditions be satisfied at μ fixed in the setting above:*

1. $\Psi(\mu)$ is nonempty and bounded.
2. F_μ^r is strongly convex on some convex open neighbourhood U of the compact set $A(\Psi(\mu))$.
3. $F_\mu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.
4. For all $x \in \Psi(\mu)$, the following set is linearly independent, where $J(x) = \{j \in \{1, \dots, \tilde{q}_1\} \mid \langle \alpha_j, x \rangle = a_j\}$:

$$\{\nabla F_\mu(Ax) \cdot A\} \cup \{\alpha_j\}_{j \in J(x)} \cup \{\beta_i\}_{i=1, \dots, \tilde{q}_2}.$$

5. $\Psi(\mu) \cap \operatorname{argmin}\{g(x) \mid x \in X\} = \emptyset$.

Then, the solution set mapping Ψ is upper Lipschitz continuous at μ in the accordingly restricted class of probability measures, i.e., there are constants $L, \delta > 0$ such that

$$\operatorname{dist}(x, \Psi(\mu)) \leq L \|F_\mu - F_\nu\|_b^{1,1} \quad \forall x \in \Psi(\nu) \forall \nu \in \mathcal{P}(\mathbb{R}^s), F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s), \|F_\mu - F_\nu\|_b^{1,1} < \delta.$$

Proof.

We are going to apply Theorem 8 with $U := \mathcal{C}_b^{1,1}(\mathbb{R}^s)$, $u_0 := F_\mu$, $q_1 := \tilde{q}_1 + 1$, $q_2 := \tilde{q}_2$, $G_1(x, u) := p - u(Ax)$, $G_j(x, u) := \langle \alpha_{j-1}, x \rangle$ ($j = 2, \dots, \tilde{q}_1 + 1$), $G_i(x, u) := \langle \beta_i, x \rangle$ ($i = 1, \dots, \tilde{q}_2$). Then, obviously, the constraint sets in Theorems 8 and 9 coincide for all $u := F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$, $\nu \in \mathcal{P}(\mathbb{R}^s)$:

$$G(x, u) \in K \iff x \in X, u(Ax) \geq p.$$

In particular, $S(u) = \Psi(\nu)$. The partial derivatives of G are calculated as

$$\nabla_x G(x, u) = \begin{pmatrix} -\nabla u(Ax)A \\ \alpha_j^T \ (j = 1, \dots, \tilde{q}_1) \\ \beta_i^T \ (i = 1, \dots, \tilde{q}_2) \end{pmatrix}, \quad \nabla_u G(x, u) = \begin{pmatrix} L \\ 0_{\tilde{q}_1 + \tilde{q}_2} \end{pmatrix},$$

where $Lu = -u(Ax)$. From the definition of $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$ one easily verifies that G belongs to the class $\mathcal{C}^{1,1}$, hence, assumption 1. of Theorem 8 is satisfied.

Next, we show:

$$\text{there is some } \hat{x} \in X \text{ such that } F_\mu(A\hat{x}) > p. \tag{6}$$

To this aim, choose some $x \in \Psi(\mu)$ according to condition 1. in our theorem. Then, $x \in X$ and $F_\mu(Ax) \geq p$. Owing to condition 4., there is a solution v of the linear system

$$\langle \nabla(F_\mu \circ A)(x), v \rangle = 1, \quad \langle \alpha_j, v \rangle = \langle \beta_i, v \rangle = 0 \ (j \in J(x), i = 1, \dots, \tilde{q}_2).$$

Then, for $\varepsilon > 0$ sufficiently small, $\hat{x} := x + \varepsilon v$ satisfies (6). Now, (6) along with condition 1. entails upper semicontinuity of Ψ at μ via Theorem 2, whence assumption 2. of Theorem

8. The quadratic growth of f required in assumption 3. of Theorem 8 was verified in the context of problem (P_ν) in [6], (Th. 8), under conditions 1., 2., 5. of our theorem together with (6). Finally, assumption 4. of Theorem 8 follows immediately from condition 4. in our theorem (recall that $\nabla_x G_1(x, u_0) = -\nabla F_\mu(Ax) \cdot A$). \square

When comparing the last Theorem with Corollary 5 which imposes the same data requirements, the stronger Lipschitz result is mainly based on two additional assumptions (leaving apart the condition 4. of linear independence in Theorem 9 which can be understood as a modification of the Slater type condition in the previous results): firstly, condition 5. requires that the chance constraint $F_\mu(Ax) \geq p$ affects the solution of the problem. If this condition is violated, no Lipschitz rate can be expected for solutions even when all remaining assumptions of Theorem 9 hold true. This can be seen from a small modification of Example 6 upon replacing the uniform distribution there by some bivariate normal distribution with independent components in order to meet the data requirements of Theorem 9. In that example, the solution set of the fixed problem with chance constraint is the same as the solution set of the unconstrained problem with removed chance constrained. As a consequence, a Hölder rate of $1/2$ results.

Secondly, the probability measures in Theorem 9 are restricted to have distribution functions in the space $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$. This applies for the fixed measure μ as well as to its perturbations ν (see statement of the result in Theorem 9) Again, without such restriction no Lipschitz rate could be obtained. We refer once more to Example 2.10 in [7] (which would have to be slightly modified in the same sense as before). In this example, all assumptions of Theorem 9 are satisfied. However the perturbed measures are just Lipschitz continuous and do not belong to $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$. They are constructed in such a way that the perturbed solution set $\Psi(\nu)$ grows at a Hölder rate of $1/2$ away from the unperturbed solution set $\Psi(\mu)$.

Although the result in Theorem 9 is stronger than that of Corollary 5 in that it improves the Hölder rate towards a Lipschitz rate, it provides only an upper estimate whereas the estimate of Corollary 5 is two-sided by relying on the Hausdorff distance. Furthermore, even the upper estimate of Theorem 9 is slightly weaker than its one-sided counterpart in Corollary 5, since, by definition of $\|\cdot\|_b^{1,1}$ and of d_K , one has

$$\|F_{\nu_1} - F_{\nu_2}\|_b^{1,1} \leq d_K(\nu_1, \nu_2) \quad \text{for all } \nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^s), F_{\nu_1}, F_{\nu_2} \in \mathcal{C}_b^{1,1}(\mathbb{R}^s).$$

Of course, imposing new restrictions raises the question of which class of probability measures still meets the new assumptions. Theorem 9 requires that both the original and all the perturbed measures have distribution functions in $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$. The following proposition identifies two classes of such measures:

Proposition 10 *Let $\nu \in \mathcal{P}(\mathbb{R}^s)$.*

1. *If ν is a nondegenerate multivariate normal distribution, then $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.*
2. *If ν is the distribution of a random vector with independent components and if the 1-dimensional distributions $\nu_i \in \mathcal{P}(\mathbb{R})$ of these components have bounded and Lipschitzian densities f_{ν_i} , then $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.*

Proof. Ad 1.: Without loss of generality, one may consider standard normal distributions (zero mean and unit variances). It is well known then (e.g. [12], p. ***), that the partial derivatives of F_ν can be calculated as

$$\frac{\partial F_\nu}{\partial x_i}(x) = \tilde{F}_{\tilde{\nu}}(\tilde{x}_i) \cdot f(x_i) \quad (i = 1, \dots, s),$$

where $\tilde{F}_{\tilde{\nu}}$ is the distribution function of some nondegenerate multivariate normal distribution $\tilde{\nu} \in \mathcal{P}(\mathbb{R}^{s-1})$, $\tilde{x}_i \in \mathbb{R}^{s-1}$ and f is the density of the 1-dimensional standard normal distribution. Taking into account that $F_\nu, \tilde{F}_{\tilde{\nu}}$ and f are bounded (say by some $M > 0$), it follows immediately that $F_\nu \in \mathcal{C}^1(\mathbb{R}^s)$ is bounded and has bounded derivative. Since $\tilde{F}_{\tilde{\nu}}$ (as a nondegenerate multivariate normal distribution function) and f are Lipschitzian on \mathbb{R}^{s-1} and \mathbb{R} , respectively, it follows that the partial derivatives of F_ν are Lipschitzian on \mathbb{R}^s (as products of functions which are bounded and Lipschitzian on \mathbb{R}^s). Hence, $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.

Ad 2.: Clearly, F_ν is bounded as a distribution function. By the assumption of independence, $F_\nu = F_{\nu_1} \cdots F_{\nu_s}$, where F_{ν_i} are the marginal distributions of ν . Since the marginal densities f_{ν_i} were assumed to be Lipschitzian, the F_{ν_i} and, hence, F_ν itself are of class \mathcal{C}^1 . The assumed boundedness of the f_{ν_i} yields that the F_{ν_i} are Lipschitzian. Furthermore,

$$\frac{\partial F_\nu}{\partial x_1} = f_{\nu_1} \cdot F_{\nu_2} \cdots F_{\nu_s}.$$

Therefore, $\frac{\partial F_\nu}{\partial x_1}$ is bounded and Lipschitzian according to the assumptions. The same argumentation applies to the other partial derivatives, whence $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$. \square

4 Illustration of the Stability Results

In this section we illustrate the obtained stability result for a simple 2-dimensional example. We consider the problem

$$\min\{x_1 + x_2 \mid P(\xi_1 \leq x_1, \xi_2 \leq x_2) \geq 1/2\},$$

where ξ is assumed to have a distribution μ which is normal with independent components of mean zero and unit variance. Clearly, this problem satisfies the basic data assumptions (BCA). The solution set of this problem consists of a singleton $\Psi(\mu) = \{q, q\}$, where $q \approx 0.55$ is the $1/\sqrt{2}$ -quantile of the 1-dimensional standard normal distribution. First, we check the assumptions of Theorem 4. Obviously, $\Psi(\mu)$ is nonempty and bounded. Next, a Slater point certainly exists, any \hat{x} with $\hat{x}_1 = \hat{x}_2 > q$ satisfies $F_\mu(\hat{x}) > F_\mu(q, q) = 1/2$. Furthermore, as μ is a normal distribution with independent components, F_μ^r is strongly concave for any $r < 0$ and on any bounded, convex set (see remarks below Theorem 4). As a consequence, F_μ^r is strongly concave on some convex, open neighbourhood of $\Psi(\mu)$. Summarizing, the first three assumptions of Theorem 4 are satisfied. Finally, in our example, g is linear (in particular: convex-quadratic), $X = \mathbb{R}^2$ is trivially polyhedral and h is linear as the identity. Hence, Corollary 5 guarantees the Hausdorff Hölder continuity with rate $1/2$ of the solution set mapping Ψ for any approximation $\nu \in \mathcal{P}(\mathbb{R}^s)$ of μ .

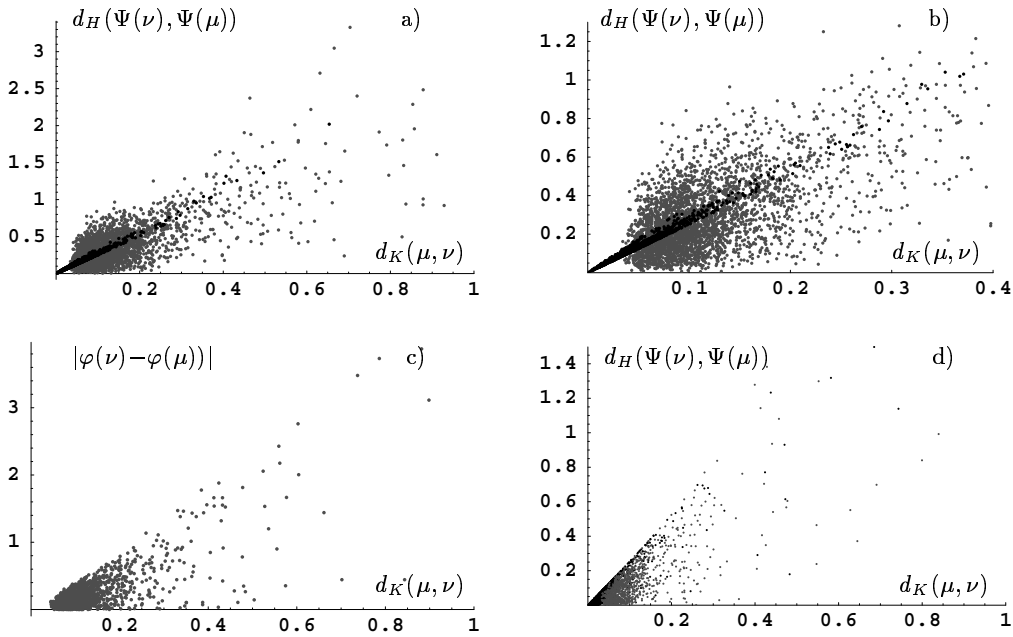


Figure 1: Illustration of stability results for simulated data

We want to focus now on two specific approximations both of which are based on a sample Z_1, \dots, Z_N of observations of ξ . The empirical measure derived from this sample is defined as $\nu = N^{-1} \sum_{i=1}^N \delta_{Z_i}$, where δ_Z is the Dirac measure placing mass one at the point Z . The empirical measure is a suitable approximation when no information at all is available about the true measure μ . If, on the other hand, partial information about μ is given, better adapted approximations may be favorable. For instance, if we know that μ in our problem is some nondegenerate multivariate normal distribution (but do not know its parameters), then a parametric approximation defining a normal distribution with mean and (co-) variances estimated from Z_1, \dots, Z_N may be useful. We want to symbolize this parametric approximation by ν' . Of course, with increasing sample size N , $d_K(\nu, \mu)$ and $d_K(\nu', \mu)$ will tend to zero in a probabilistic sense, and $d_K(\nu', \mu)$ will do so even faster than $d_K(\nu, \mu)$. The issue we want to address here is convergence of the approximating solution sets, i.e., dependence of $d_H(\Psi(\nu), \Psi(\mu))$ on $d_K(\nu, \mu)$. To this aim, several thousand samples of ξ were simulated according to its distribution μ . The sample size varied up to a few hundred.

Figure 1 a) illustrates the results for the parametric (black dots) and empirical (gray dots) approximations. Clearly, in both cases the approximating solutions converge to the true solution when the approximating measure converges to the true measure. Indeed, this kind of qualitative stability is already ensured by the first two assumptions of Theorem 4 via Theorem 2. From a quantitative point of view, however, the solution sets of parametric approximations seem to converge much faster (in the worst case) than those of empiric approximation. This is particularly obvious in a region close to the origin which has

been magnified in Figure 1 b). According to the diagram, there is no doubt that there exists an upper Lipschitz estimation for the parametric approximation, whereas in case of the empiric estimation increasingly large ratios between the two distances seem to be possible when $d_K(\nu, \mu)$ tends to zero. This suggests a Non-Lipschitzian relation. At least, Corollary 5 guarantees that the corresponding cloud of points lies below some function $\alpha\sqrt{d_K(\nu, \mu)}$, where $\alpha > 0$ is sufficiently large.

As far as the parametric approximation is concerned, its Lipschitzian behavior is supported by Theorem 9. To see this, recall that both the original and the approximating measures are normal distributions, hence, their distribution functions belong to the space $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$ according to Proposition 10. Furthermore, the gradient of a (nondegenerate) normal distribution function is always nonzero which yields condition 4. of Theorem 9. Finally, owing to the fact that the objective function in our example is linear, condition 5. of Theorem 9 is trivially fulfilled. It has to be noted, that Theorem 9 provides a Lipschitz result with respect to the distance $\|F_\mu - F_\nu\|_b^{1,1}$, whereas Figure 1 b) even suggests a Lipschitzian relation with respect to the stronger Kolmogorov distance $d_K(\mu, \nu)$.

As far as optimal values are concerned, Theorem 2 guarantees a Lipschitzian estimation for any approximating measure. This is observed empirically in Figure 1 c) for the example of empirical approximation (the better behaved parametric case is omitted here).

Finally, we may reduce our example to a 1-dimensional setting, i.e., to the problem

$$\min\{x|P(\xi \leq x) \geq 1/2\},$$

where ξ is assumed to have a standard normal distribution μ . In this situation, the dependence of Hausdorff distances between solution sets on Kolmogorov distances between measures is seen from Figure 1 d) to be of Lipschitzian nature for both types of approximations (gray dots on top of black dots). Again, this observation is supported by our results via Theorem 7, according to which the Lipschitz rate results for any approximating measure.

5 Appendix

Proposition 11 *For $r < 0$ and $\nu \in \mathcal{P}(\mathbb{R}^s)$ it holds: If $F_\nu(y) \geq w > 0$ for all $y \in Q \subseteq \mathbb{R}^s$, then there exist constants $c, \delta > 0$ such that*

$$|F_\nu^r(y) - F_{\nu'}^r(y)| \leq cd_K(\nu, \nu') \quad \forall y \in Q \quad \forall \nu' \in \mathcal{P}(\mathbb{R}^s), d_K(\nu, \nu') < \delta.$$

Proof. Note that

$$|u^r - v^r| \leq |r| \max\{u^{r-1}, v^{r-1}\} |u - v| \quad \forall u, v > 0.$$

Then, choosing $\delta := w/2$, one has

$$F_{\nu'}^r(y) \geq w/2 > 0 \quad \forall y \in Q \quad \forall \nu' \in \mathcal{P}(\mathbb{R}^s), d_K(\nu, \nu') < \delta.$$

Fix c as $|r|(w/2)^{r-1}$. □

Proposition 12 *With the assumptions and notations of Lemma 3 assume that*

1. *Y is Hausdorff Hölder continuous with rate $1/2$ at μ , i.e., there are constants $\rho, \delta > 0$ such that*

$$d_H(Y(\mu), Y(\nu)) \leq \rho d_K^{1/2}(\mu, \nu) \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \delta.$$

2. *σ is Hausdorff Hölder continuous with rate κ^{-1} on Y_V , i.e., there exists $L > 0$ such that*

$$d_H(\sigma(z), \sigma(y)) \leq L d^{\kappa^{-1}}(y, z) \quad \forall z, y \in Y_V.$$

Then, Ψ is Hausdorff Hölder continuous with rate $(2\kappa)^{-1}$ at μ . More precisely, it holds that

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L \rho^{\kappa^{-1}} [d_K(\mu, \nu)]^{(2\kappa)^{-1}} \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \delta.$$

Proof. For a nonempty and closed subset $Q \subseteq \mathbb{R}^s$ and $y \in \mathbb{R}^s$ denote by $\text{proj}^Q(y)$ the projection of y onto Q . Note that for $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \delta$ and small enough δ , one has $\Psi(\nu) \neq \emptyset$ (Theorem 2) and $Y(\nu) \neq \emptyset$ by (5). Furthermore, the sets $Y(\nu)$ are closed (see proof of statement 4. in Lemma 3 provided in [5], Lemma 1). Consequently, proj applies to these sets $Y(\nu)$. Recalling that $Y(\nu) \subseteq Y_V$, it follows from the assumptions and from (5), that for $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \delta$

$$\begin{aligned} d_H(\Psi(\mu), \Psi(\nu)) &= \max\left\{ \sup_{x \in \Psi(\mu)} d(x, \Psi(\nu)), \sup_{x' \in \Psi(\nu)} d(x', \Psi(\mu)) \right\} \\ &= \max\left\{ \sup_{y \in Y(\mu)} \sup_{\alpha \in \sigma(y)} d(\alpha, \sigma(Y(\nu))), \sup_{y' \in Y(\nu)} \sup_{\beta \in \sigma(y')} d(\beta, \sigma(Y(\mu))) \right\} \\ &\leq \max\left\{ \sup_{y \in Y(\mu)} \sup_{\alpha \in \sigma(y)} d(\alpha, \sigma(\text{proj}^{Y(\nu)}(y))), \right. \\ &\quad \left. \sup_{y' \in Y(\nu)} \sup_{\beta \in \sigma(y')} d(\beta, \sigma(\text{proj}^{Y(\mu)}(y'))) \right\} \\ &\leq L \max\left\{ \sup_{y \in Y(\mu)} d^{\kappa^{-1}}(y, \text{proj}^{Y(\nu)}(y)), \sup_{y' \in Y(\nu)} d^{\kappa^{-1}}(y', \text{proj}^{Y(\mu)}(y')) \right\} \\ &\leq L \left[\max\left\{ \sup_{y \in Y(\mu)} d(y, Y(\nu)), \sup_{y' \in Y(\nu)} d(y', Y(\mu)) \right\} \right]^{\kappa^{-1}} \\ &\leq L [d_H(Y(\mu), Y(\nu))]^{\kappa^{-1}} \\ &\leq L \rho^{\kappa^{-1}} [d_K(\mu, \nu)]^{(2\kappa)^{-1}}. \end{aligned}$$

□

Lemma 13 *Under the assumptions of Theorem 2 and with the notations of Lemma 3, consider the parametric program*

$$(\tilde{P}_\nu) \quad \min \{ \pi(y) \mid y \in Y_V, F_\nu(y) \geq p \} \quad (\nu \in \mathcal{P}(\mathbb{R}^s))$$

near $\mu \in \mathcal{P}(\mathbb{R}^s)$, the solution set mapping and optimal value function of which are given by Y and φ , respectively (see Lemma 3). Let the following assumption be satisfied in addition, where $r < 0$ refers to the exponent of concavity of μ

- F_μ^r is strongly convex on some convex open neighbourhood U of $Y(\mu)$.

Then, Y is Hausdorff Hölder continuous with rate $1/2$ at μ .

Proof. Setting $b_\nu(y) := F_\nu^r(y) - p^r$ for $\nu \in \mathcal{P}(\mathbb{R}^s)$, the original problem (\tilde{P}_μ) may be written as

$$(\tilde{P}_\mu) \quad \min \{ \pi(y) \mid y \in Y_V, b_\mu(y) \leq 0 \}.$$

As a consequence of the r -concavity of μ (where $r < 0$), F_μ^r is a convex (possibly extended-valued) function. Therefore, b_μ is a convex function finite-valued on Y_V (see definition of Y_V). Then, in view of 1. and 2. in Lemma 3, (\tilde{P}_μ) is a convex program which satisfies the Slater condition $b_\mu(\hat{y}) < 0$ for some $\hat{y} \in Y_V$. Indeed, we may choose $x^* \in \Psi(\mu) \neq \emptyset$ (first assumption of Th. 2), hence $x^* \in X \cap V$ and $b_\mu(h(x^*)) \leq 0$. Furthermore, $\hat{x} \in X$ taken from the second assumption of Theorem 2 satisfies $b_\mu(h(\hat{x})) < 0$. With F_μ^r being nondecreasing as a distribution function, the composition $F_\mu^r \circ h$ is convex too due to F_μ^r being nonincreasing ($r < 0$) and to h having concave components. Therefore, $b_\mu \circ h$ is convex and, for sufficiently small $\lambda > 0$, $x_\lambda := \lambda \hat{x} + (1 - \lambda)x^*$ satisfies $b_\mu(h(x_\lambda)) < 0$. Now, one may take $\hat{y} := h(x_\lambda)$.

Statement 4. in Lemma 3 and Proposition 11 guarantee that for some $c, \delta' > 0$

$$Y(\nu) \subseteq U, |b_\nu(y) - b_\mu(y)| \leq c d_K(\mu, \nu) \quad \forall y \in Y_V, \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \delta'. \quad (7)$$

Finally, the additional assumption on strong convexity of F_μ^r on U means in particular that

$$b_\mu(y_1/2 + y_2/2) \leq b_\mu(y_1)/2 + b_\mu(y_2)/2 - \rho \|y_1 - y_2\|^2 \quad \forall y_1, y_2 \in U \quad (8)$$

for some $\rho > 0$. We proceed by case distinction with respect to the relation between $Y(\mu)$ and the solution set $Q := \arg \min \{ \pi(y) \mid y \in Y_V \}$ of the relaxed problem (\tilde{P}_μ) where the chance constraint $b_\mu(y) \leq 0$ is omitted.

case 1: $Y(\mu) \cap Q = \emptyset$.

Choose some $y^* \in Y(\mu)$ (recall that $Y(\mu) \neq \emptyset$ due to $\Psi(\mu) \neq \emptyset$ and to (5)). Since π and b_μ are finite-valued on Y_V (statement 2. of Lemma 3 and $\varphi(\mu) = \pi(y^*) > -\infty$ (see (4))), the Slater condition shown above for problem (\tilde{P}_μ) ensures the existence of a Lagrange multiplier $\lambda^* \geq 0$ such that (cf. [13], Cor. 28.2.1)

$$\pi(y^*) = \min \{ \pi(y) + \lambda^* b_\mu(y) \mid y \in Y_V \} \text{ and } \lambda^* b_\mu(y^*) = 0. \quad (9)$$

By the case 1- assumption, one has $\lambda^* > 0$ and so $\pi + \lambda^* b_\mu$ is strongly convex on $Y_V \cap U$ due to the additional assumption in this lemma. This implies

$$\tilde{\rho} \|y - y^*\|^2 \leq \pi(y) + \lambda^* b_\mu(y) - \pi(y^*) \quad \text{for all } y \in Y_V \cap U. \quad (10)$$

for some $\tilde{\rho} > 0$ (due to $\lambda^* b_\mu(y^*) = 0$ and y^* being a minimizer in (9)). In particular, y^* is the unique minimizer of (\tilde{P}_μ) , i.e., $Y(\mu) = \{y^*\}$. For an arbitrary ν taken from (7), (10)

applies. Using the results of Lemma 3 and the fact that $b_\nu(y) \leq 0$ for all $y \in Y(\nu)$ one arrives at the asserted Hölder continuity with respect to the Hausdorff distance:

$$\begin{aligned}
d_H(Y(\mu), Y(\nu)) &= \sup_{y \in Y(\nu)} d(y, y^*) \leq \tilde{\rho}^{-1/2} \sup_{y \in Y(\nu)} [\pi(y) - \pi(y^*) + \lambda^*(F_\mu^r(y) - p^r)]^{1/2} \\
&\leq \tilde{\rho}^{-1/2} \sup_{y \in Y(\nu)} [\varphi(\nu) - \varphi(\mu) + \lambda^*(b_\mu(y) - b_\nu(y))]^{1/2} \\
&\leq \tilde{\rho}^{-1/2} [Ld_K(\mu, \nu) + \lambda^*cd_K(\mu, \nu)]^{1/2} \\
&\leq \tilde{\rho}^{-1/2}(L + \lambda^*c)^{1/2}d_K(\mu, \nu)^{1/2} \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s), d_K(\mu, \nu) < \min\{\delta', \delta\}
\end{aligned}$$

with $L, \delta > 0$ from Theorem 2.

case 2: $Y(\mu) \cap Q \neq \emptyset$.

In this case, $Y(\mu)$ has the simple representation

$$Y(\mu) = \{y \in Q \mid b_\mu(y) \leq 0\}. \quad (11)$$

Note that Q is closed and convex by the properties of π and Y_V stated in Lemma 3.

case 2.1 $\exists \bar{y} \in Y(\mu), b_\mu(\bar{y}) < 0$.

Then, \bar{y} is a Slater point of the constraint $b_\mu(y) \leq 0$ with respect to Q . As a consequence of results in [14] (Theorem 3.2, Corollary 3.7 and Lemma A.2), each $y \in Y(\mu)$ is supplied with neighbourhoods V_y of y and U_y of μ such that for moduli $L_y > 0$

$$\begin{aligned}
d(y', Y(\nu)) &\leq L_y d_K(\mu, \nu) \quad \forall \nu \in U_y \quad \forall y' \in Y(\mu) \cap V_y \\
d(y', Y(\mu)) &\leq L_y d_K(\mu, \nu) \quad \forall \nu \in U_y \quad \forall y' \in Y(\nu) \cap V_y
\end{aligned}$$

The compactness of $Y(\mu) \subseteq Y_V$ (statement 1. of Lemma 3) then allows to extract a neighbourhood \tilde{U} of μ and an open set \tilde{V} containing $Y(\mu)$ such that

$$\begin{aligned}
d(y, Y(\nu)) &\leq L d_K(\mu, \nu) \quad \forall \nu \in \tilde{U} \quad \forall y \in Y(\mu) \\
d(y, Y(\mu)) &\leq L d_K(\mu, \nu) \quad \forall \nu \in \tilde{U} \quad \forall y \in Y(\nu) \cap \tilde{V}
\end{aligned}$$

with some $L > 0$. By upper semicontinuity of Y (statement 4. of Lemma 3), one has $Y(\nu) \subseteq \tilde{V}$ for all $\nu \in \mathcal{P}(\mathbb{R}^s)$, $d_K(\mu, \nu) < \tilde{\delta}$ with some $\tilde{\delta} > 0$. Hence, even Hausdorff Lipschitz continuity of Y at μ follows from the above inequalities: $d_H(Y(\mu), Y(\nu)) \leq Ld_K(\mu, \nu)$ for all $\nu \in \mathcal{P}(\mathbb{R}^s)$, $d_K(\mu, \nu) < \delta^*$ with some $\delta^* > 0$. This, of course, implies the asserted Hölder continuity with rate $1/2$.

case 2.2 $b_\mu(y) = 0 \forall y \in Y(\mu)$.

The convexity of $Y(\mu)$ along with (8) yield that $Y(\mu)$ reduces to a singleton, say $Y(\mu) = \{y^*\}$. Then, $b_\mu(y^*) = 0$ and $y^* \in Q \subseteq Y_V$ by (11). For any ν satisfying (7), let $y \in Y(\nu) \subseteq U$ be arbitrary, hence $y \in Y_V$ and $b_\nu(y) \leq 0$. Put

$$\lambda' := \inf\{\lambda \geq 0 \mid b_\mu(\lambda y^* + (1 - \lambda)y) \leq 0\}.$$

Then, $\lambda' \in [0, 1]$. Define $y' := \lambda' y^* + (1 - \lambda')y$. Assume first that $\lambda' > 0$. Since the convex function $\alpha(\lambda) = b_\mu(\lambda y^* + (1 - \lambda)y)$ is upper semicontinuous on $[0, 1]$ and continuous on $(0, 1)$, it follows that $b_\mu(y') = 0$. Since, for $\lambda' > 0$, $b_\mu(y) > 0$, one has $y' \neq y$ and $b_\mu(y/2 + y'/2) > 0$ according to the definition of y' . Then, (7) and (8) yield

$$cd_K(\mu, \nu) \geq b_\mu(y) - b_\nu(y) \geq b_\mu(y)/2 + b_\mu(y')/2 \geq b_\mu(y/2 + y'/2) + \rho \|y - y'\|^2 \geq \rho \|y - y'\|^2,$$

whence

$$\|y - y'\| \leq \sqrt{c/\rho} \sqrt{d_K(\mu, \nu)}. \quad (12)$$

In the excluded case of $\lambda' = 0$, the same inequality follows trivially from $y' = y$. Now, we want to estimate the distance between y' and y^* , hence, without loss of generality, we may assume that $y' \neq y^*$. Then, $\lambda' < 1$ and $y' \notin Q$ (if $y' \in Q$, then $y' \in Y(\mu)$ due to $b_\mu(y') = 0$ and (11), whence a contradiction to $Y(\mu) = \{y^*\}$). Now, $y \notin Q$ since $y^* \in Q$ and Q is convex (otherwise the contradiction $y' \in Q$). Consequently, $\pi(y) > \pi(y^*)$. Put, $y'' := y^*/2 + y'/2$, hence $y'' = \frac{\lambda'+1}{2}y^* + \frac{1-\lambda'}{2}y$, which is a convex combination of y^* and y . Then, $y'' \in Y_V \cap U$ due to convexity of $Y_V \cap U$. It follows that

$$\pi(y'') \leq \frac{\lambda' + 1}{2} \pi(y^*) + \frac{1 - \lambda'}{2} \pi(y) < \pi(y).$$

If $b_\nu(y'') \leq 0$, then a contradiction to $y \in \Psi(\nu)$ results, hence $b_\nu(y'') > 0$. Again referring to (7) and (8), it follows that

$$\begin{aligned} cd_K(\mu, \nu) \geq b_\nu(y'') - b_\mu(y'') &\geq -b_\mu(y^*/2 + y'/2) \geq -(b_\mu(y^*) + b_\mu(y'))/2 + \rho \|y^* - y'\|^2 \\ &= \rho \|y^* - y'\|^2. \end{aligned}$$

Combining this with (12), one arrives at the desired estimation

$$d_H(Y(\mu), Y(\nu)) = \sup_{y \in Y(\nu)} \|y - y^*\| \leq 2\sqrt{c/\rho} \sqrt{d_K(\mu, \nu)}.$$

□

In the case of a 1-dimensional random variable, the assertion of the previous lemma can be sharpened even without the strong convexity assumption made there:

Lemma 14 *If $s = 1$ then, under the assumptions of Theorem 2, Y is Hausdorff Lipschitz continuous (i.e., Hausdorff Hölder continuous with rate $\kappa = 1$) at μ .*

Proof. We consider the parametric program from Lemma 13 which Y is the solution mapping of:

$$(\tilde{P}_\nu) \quad \min \{ \pi(y) \mid y \in Y_V, F_\nu(y) \geq p \} \quad (\nu \in \mathcal{P}(\mathbb{R}))$$

We have $Y_V = [a, b]$ for some $a, b \in \mathbb{R}$ (see 1. in Lemma 3). Choosing some $x^* \in \Psi(\mu) \subseteq X$ according to the assumption of Theorem 2, it follows that $h(x^*) \in Y_V \neq \emptyset$, hence $a \leq b$. Since F_ν is upper semicontinuous and nondecreasing as a distribution function, one gets

$$\{y \in \mathbb{R} \mid F_\nu(y) \geq p\} = [\alpha(\nu), \infty), \quad \alpha(\nu) := \min\{y \in \mathbb{R} \mid F_\nu(y) \geq p\}.$$

Clearly, $\{\alpha(\nu)\}$ is the solution set of a parametric program of type (P_ν) (see introduction) which at the fixed measure μ satisfies the basic data assumptions (BCA) (with $g(x) = h(x) = x$ and $X = \mathbb{R}$). Since $p \in (0, 1)$ and F_μ is a distribution function, there exists some $\bar{y} \in \mathbb{R}$ with $F_\mu(\bar{y}) > p$. Now, Theorem 2 allows to derive the existence of $L, \delta > 0$ such that

$$|\alpha(\nu) - \alpha(\mu)| = |\varphi(\nu) - \varphi(\mu)| \leq Ld_K(\mu, \nu) \quad \forall \nu \in \mathcal{P}(\mathbb{R}), d_K(\mu, \nu) < \delta,$$

where $\varphi(\nu)$ refers to the optimal value function of the parametric problem defining $\alpha(\nu)$. Summarizing, we may rewrite (\tilde{P}_ν) as

$$(\tilde{P}_\nu) \quad \min \{ \pi(y) \mid y \in [b(\nu), b] \} \quad (\nu \in \mathcal{P}(\mathbb{R})),$$

where $b(\nu) := \max\{\alpha(\nu), a\}$ satisfies

$$|b(\nu) - b(\mu)| \leq Ld_K(\mu, \nu) \quad \forall \nu \in \mathcal{P}(\mathbb{R}), d_K(\mu, \nu) < \delta. \quad (13)$$

We argue that $b(\nu) \leq b$ for all $\nu \in \mathcal{P}(\mathbb{R})$ with $d_K(\mu, \nu) < \tilde{\delta}$ and some $\tilde{\delta} > 0$. This is obvious from (13) if $b(\mu) < b$. If $b(\mu) = b$, then we refer to some $\hat{y} \in Y_V$ with $F_\mu(\hat{y}) > p$ (see proof of Lemma 13). Consequently, $a = b = \hat{y}$ and $F_\mu(b) > p$. Then, $F_\nu(b) \geq p$ and, hence, $b(\nu) \leq b$ for all $\nu \in \mathcal{P}(\mathbb{R})$ with $d_K(\mu, \nu) < \tilde{\delta} := F_\mu(b) - p$.

Now, π is a lower semicontinuous, convex and finite function on the nonempty intervals $[b(\nu), b] \subseteq Y_V$ (see Lemma 3). In particular, $Y(\nu) \neq \emptyset$ for all $\nu \in \mathcal{P}(\mathbb{R})$ with $d_K(\mu, \nu) < \tilde{\delta}$. Elementary calculus shows that

$$d_H(Y(\nu), Y(\mu)) \leq |b(\nu) - b(\mu)| \quad \forall \nu \in \mathcal{P}(\mathbb{R}), d_K(\mu, \nu) < \tilde{\delta}.$$

Along with (13), this yields the assertion of the Lemma. \square

References

- [1] J.F. Bonnans and A. Shapiro, Nondegeneracy and quantitative stability of parametrized optimization problems with multiple solutions, *SIAM Journal on Optimization* 8 (1998), 940–946.
- [2] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
- [3] C. Borell, Convex Sets in d-Space, *Periodica Mathematica Hungarica* 6 (1975), 111–136.
- [4] N. Gröwe, Estimated stochastic programs with chance constraints, *European Journal of Operational Research* 101 (1997), 285–305.
- [5] R. Henrion, Qualitative stability of convex programs with probabilistic constraints, in: (V.H. Nguyen, J.-J. Strodiot and P. Tossings eds.) *Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 481, Springer, Berlin 2000, 164–180.

- [6] R. Henrion and W. Römisch, Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Mathematical Programming* 84 (1999), 55-88.
- [7] R. Henrion and W. Römisch, Stability of solutions to chance constrained stochastic programs, in: (J. Guddat, R. Hirabayashi, H.Th. Jongen and F. T wilt eds.) *Parametric Optimization and Related Topics V*, Peter Lang, Frankfurt a.M. 2000, 95-114.
- [8] R. Henrion and W. Römisch, in preparation.
- [9] V. Kaňková, A note on multifunctions in stochastic programming, in: *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer, Berlin 1998, 154-168.
- [10] D. Klatté and G. Thiere, Error bounds for solutions of linear equations and inequalities, *ZOR - Mathematical Methods of Operations Research* 41 (1995) 191-214.
- [11] R. Lepp, Discrete approximation of extremum problems with chance constraints, in: *Stochastic Optimization Techniques* (K. Marti ed.), Lecture Notes in Economics and Mathematical Systems Vol. 513, Springer, Berlin 2000, 21-33.
- [12] A. Prekopá, *Stochastic Programming*, Kluwer, Dordrecht, 1995.
- [13] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
- [14] W. Römisch and R. Schultz, Stability analysis for stochastic programs, *Annals of Operations Research* 30 (1991) 241-266.
- [15] J. Wang, Continuity of feasible solution sets of probabilistic constrained programs, *Journal of Optimization Theory and Applications* 63 (1989), 79-89.
- [16] R.-J.-B. Wets, Stochastic programs with chance constraints: Generalized convexity and approximation issues, in: *Generalized Convexity, Generalized Monotonicity: Recent Results* (J.-P. Crouzeix, J.-E. Martínez-Legaz and M. Volle eds.), Kluwer, Dordrecht 1998, 61-74.